

THE PRINCIPAL COMPONENTS OF
RESPONSE STRENGTH

PETER R. KILLEEN AND SCOTT S. HALL

ARIZONA STATE UNIVERSITY

As Skinner (1938) described it, response strength is the “state of the reflex with respect to all its static properties” (p. 15), which include response rate, latency, probability, and persistence. The relations of those measures to one another was analyzed by probabilistically reinforcing, satiating, and extinguishing pigeons’ key pecking in a trials paradigm. Reinforcement was scheduled according to variable-interval, variable-ratio, and fixed-interval contingencies. Principal components analysis permitted description in terms of a single latent variable, *strength*, and this was validated with confirmatory factor analyses. Overall response rate was an excellent predictor of this state variable.

Key words: factor analysis, latency, persistence, response probability, response rate, strength, pigeons

What is the proper dependent variable in the experimental analysis of behavior, and how should it be measured? We start at the beginning, with Skinner (1938):

The following aspects of the system bear upon the problem of the measure to be taken: (1) the definition of behavior as that part of the activity of the organism which affects the external world; (2) the practical isolation of a unit of behavior; (3) the definition of a response as a class of events; and (4) the demonstration that the rate of responding is the principal measure of the strength of an operant. It follows that the main datum to be measured in the study of the dynamic laws of an operant is the length of time elapsing between a response and the response immediately preceding it or, in other words, the rate of responding. (p. 58)

At this point in Skinner’s theoretical development, rate was important not in itself, but as the principal component of “strength.” The construct of *strength* “describes the state of the reflex with respect to all its static properties at once” (1938, p. 15). There are many properties of a response that are affected by conditioning (e.g., force, rate, persistence, probability, topography, etc.), and if these covary, it is reasonable to assume that a common “thing” is conditioned, of which these

are by-products. The choice of response rate (Point 4) follows from the adoption of the free-operant procedure: Absent an eliciting stimulus, the static properties of Pavlovian reflexes—threshold, latency, and magnitude—are unavailable. Response strength was Skinner’s chief intervening variable: “The conditioning here is again a matter of a change in strength” (p. 21) and “the strength of a reflex is proportional to its reserve” (p. 27).

Hypothetical Constructs

What is the nature of Skinner’s construct, *strength*? It is the measure of the *operant*; both are hypothetical constructs. A distinction is often made between intervening variables (InV) and hypothetical constructs (HC). Intervening variables are placeholders in models. Physical momentum is an example: It is the product of mass and velocity, nothing more. Hypothetical constructs correspond to physical entities or processes. Mass and oxidation are examples. With sufficient converging evidence, HCs graduate to “facts.” Gravity started life contentiously as an HC, but is now considered a fact. Gravity waves remain an HC. Biological evolution is a fact, with much of the evidence for it gathered by Victorian naturalists. Theories are organized systems of facts, hypotheses, and rules of inference. Laymen often confuse theories such as Darwin’s or the new synthesis with mere hypotheses. Hypotheses, which often involve HCs, are elements of theories.

Both InVs and HCs are useful tools, and are often difficult to discriminate from one another. Both are HCs in the literal sense,

This research was supported by NSF Grant IBN 9408022 and NIMH Grant K05 MH01293. Tony Nevin’s comments greatly improved prior drafts. Scott Hall is now at the School of Psychology, University of Birmingham, England.

Address correspondence to Peter Killeen, Department of Psychology, Box 1104, Arizona State University, Tempe, Arizona 85287-1104 (E-mail: killeen@asu.edu).

with HCs having claims for existential status and InVs not. InVs are often incorrectly reified: A fast train is said to have “a lot of momentum,” when all it has is significant mass moving at appreciable velocity. Zuriff (1985) devotes a chapter to a thoughtful explication of these concepts, their philosophical justification, and strategic utility. The *operant* and its measure, *strength*, are InVs. Throughout this paper, however, the more general sense of HC is used, because the presence or absence of existential claims for such constructs adds nothing useful to the analyses. The general category is thus the HC, with InVs a proper subset.

The operant response is a member of a class (Point 3). Any particular response is a *fait accompli*. To talk of a change in the rate of a response requires a multiplicity of responses that are sufficiently “alike” to be counted as instances of the same class. How do they get into the class? How alike must two responses be to be members, and alike along what dimensions? “An operant regarded as a functional part of behavior [is] defined at levels of specification marked by the orderliness of dynamic changes” (Skinner, 1938, p. 40). This is pragmatic: Good definitions are those that provide good data. Dynamic changes result from motivational operations such as deprivation and extinction. If responses to a red key are extinguished and responses to a green key also decrease in frequency, then they are members of the same class (Killeen, 1988). Skinner also defined *alike* in terms of having the same effect on the environment (Point 1)—typically the moving of the same switch a minimal distance (Point 2). This definition is also pragmatic—good definitions permit convenient measurement—although some important activities “which affect the external world” (p. 58) may be difficult to measure, and thus may be neglected. For better or worse, easy measures may drive difficult measures out of circulation.

The necessity of the HC of strength. Skinner’s (1938) measure of the operant, *strength*, played the same theoretical role as Hull’s *reaction potential* (Amsel & Rashotte, 1984; Hull, 1943); it is a provisional identification of a thing that different instances of a class (an operant) have in common that make them members of the class. The covariation of red- and green-key responding over dynamic

changes in the red-key operant make them members of the same class. But covariation of what aspect of the response? Strength.

One way of gauging strength is by measuring response rate. But that is not always possible. Turning off the keylight does not reduce strength to zero. The rates of marrying, receiving the doctorate, and paying taxes reveal nothing about the motivation for those responses. This does not mean that they have no strength, nor does it rule out measurement in terms of latency or other properties. Without a construct such as strength, however, there would be no justification for assuming that latency or probability or rate might all be indicators of the same underlying state variable. But the construct of strength itself has not received adequate justification. That is the goal of this paper.

For every successful construct such as gravity, there is a mistake such as caloric. To minimize such false alarms, Skinner (1938) required that descriptions be kept at the same epistemological level as the things described. This ultraconservative strategy is problematic for his theory of behavior. Covariation among responses or properties cannot then be taken as evidence of an underlying state, strength, which identifies them as members of the same class. It is as though Lavoisier noted the cooccurrence of smoke, heat, light, and changes in weight, but refrained from hypothesizing an underlying process such as oxidation, because that was at a different level than the things described.

The parsimony of the HC of strength. Skinner (1938) observed that invoking hidden states could make a theory more parsimonious. This occurs when many independent manipulations effect similar changes in those states, and many dependent variables—say, rate, latency, probability, and preference—are changed in predictable ways by changes in that state. The argument was clearly developed by N. E. Miller (1959). If there are five independent variables and five dependent variables, rather than identify 25 linkages from each, we may only need 10: five from each independent variable to the state variable and five from it to the dependent variables. Constructs such as strength or drive or reinforcement are antiparsimonious if one confines attention to variations in one independent variable and measurement of one

dependent variable. This recognition was manifest in the debates over the circularity of the law of effect. Constructs are neutral with two independent variables and two dependent variables, and beyond that, they are parsimonious. It is therefore intellectually conservative to identify such unifying constructs wherever possible. This assumes the independent variables, or simple transformations of them, effect proportional changes in the dependent variables, or in simple transformations to them. As more complicated transformations are introduced, degrees of freedom again become compromised.

Most work in the experimental analysis of behavior studies the effects of various independent variables on one or another dependent variable. The purpose of the present article is to define the transformations that relate popular dependent variables to the central construct, strength.

Objections to general processes. Emphasis on idiosyncrasies of particular responses—the “constraints on learning” literature (e.g., Domjan & Galef, 1983; Timberlake & Lucas, 1990), may be seen as raising impediments to the concept of strength, but this is not the case. Garcia, McGowan, and Green (1972) correctly objected to the ubiquitous use of a lever or key, and to the assumption that all responses and all responders behave alike. But that assumption is not a bad place to start. If no responses are alike, nomothetic science is impossible; if it is assumed that all are alike and some types of responses are found to function differently than others, a new class of responses becomes necessary. Prepared, neutral, and contraprepared responses (Seligman, 1970) are, de facto, members of different types of operant classes. Their equations of motion are different. No problem: This is a feature, not a bug. It is the way science progresses. Hypothetical constructs provide the basis for a unified theory of behavior despite different operating characteristics for different responses. Rats learn to flee signals of shock more quickly than they learn to lever press to avoid those shocks. Here the motivational operations are the same; what is necessary is a way of measuring the differential effect of those operations on the different response classes. This is our task.

The Properties of a Response

The construct of *strength* “describes the state of the reflex with respect to all its static properties at once” (Skinner, 1938, p. 15). Which properties? Our tactic, in the empirical spirit of Skinner’s generic definition of the stimulus and response, is to include all properties that can be shown to covary. The analysis presented in this paper is limited to correlated properties of a single class of operant responses, with the analysis of choice deferred to a subsequent paper.

Response magnitude. A candidate static property of the reflex is the magnitude of the response (R) relative to that of the stimulus (S) (the R–S ratio). Operant responses show less covariation of response and reinforcer magnitudes: Rats do pull harder away from a large shock than a small one (Brown, 1948; N. E. Miller, 1959; Townsend & Busemeyer, 1989). But response magnitude is easily affected by the topography of the unconditioned response (Allan, 1992; Ploog & Zeigler, 1996). There are instances in which, even when contingencies are held constant, response force may covary negatively with other measures of strength (Amsel, 1962; Logan, 1956; Notterman & Block, 1960; Svartdal, 1993).

Counterintuitively perhaps, response amplitude (Herrick, 1963; Kellicutt, 1967) and duration are larger and longer for responding during extinction, and when under the control of a negative discriminative stimulus (S–) than under control of a positive discriminative stimulus (S+) (Margulies, 1961). This is also true of response force, which increases in extinction (Levine & Loesch, 1967) and is negatively correlated with the probability (Mattes, Ulrich, & Miller, 1997) and rate (Filion, Fowler, & Notterman, 1969) of responding. Furthermore, response duration may show faster differentiation under discrimination training than does response rate (Herrick, 1963). Significant negative covariation would be just as useful and important as positive covariation. Kimble (1961), however, presented a table of intercorrelations for various measures of respondent conditioning, and found them “too low to support a common-process view” (p. 111; also see Mackintosh, 1974, pp. 65, 144). With the possible exception of extinction bursts (Amsel, 1962),

the evidence for a consistent relation between magnitude and duration of responses on the one hand and motivational and conditioning operations on the other is equivocal. Magnitude will not be included in this analysis.

Rate. Rates of responding and rates of reinforcement are *molar* variables (Baum, 1973; Rachlin, 1994), and as such require specification of time (or response) windows within which they are to be measured, decisions about weighting of events within those windows (all equally, or weighting the most recent most heavily, and so on), and decisions about the nature of the average: average interresponse times (IRTs) or average rates (the reciprocals of IRTs). In the excerpt cited above, Skinner conflated rate with IRTs; but rates (1/IRTs) are not “other words” for the time between responses. The distinction is important when considering how to average the constituent events. Rate is typically measured as the reciprocal of the average (arithmetic mean) of the IRTs $[1/(\sum \text{IRT}_i/n)]$. This is equivalent to the number of responses in the interval divided by the sum of IRTs in the interval. The estimate of rate from a single IRT is its reciprocal; and the average of the reciprocals of a set of IRTs is their harmonic mean:

$$\left(\sum_{i=1}^n \frac{1}{\text{RT}_i} \right) / n.$$

These are not the same: The harmonic mean weights short IRTs more heavily than long ones, whereas the arithmetic mean weights long IRTs more heavily than short ones. Only the geometric mean, $\text{antilog}[\sum \log(\text{IRT}_i)/n]$, weights all IRTs equally.¹ To anticipate our results, we find that these later alternative av-

erages offer no advantage over the traditional measure of rates (viz. the reciprocal of the average IRT).

Latency. This is the time between one event and another. If the first event is the onset of an opportunity to respond and the second is a response, the epoch is called a *latency* or *response time*. These are the chief dependent variables of cognitive psychologists (see Meyer, Osman, Irwin, & Yantis, 1988, for a synopsis and “family tree of mental chronometry”). If the first event is a response, then the epoch is called an interresponse time. If it is a reinforcer, it is often called a postreinforcement pause.

As is the case with IRTs, which may be thought of as response-response latencies, there are various ways to average latencies. The average of the times is most common. Also possible is a “speed” analogue, calculated as the average of the reciprocals of the times between trial onset and the response. Finally, there is the geometric mean of the latencies, whose weighting is equal for all events. Luce (1986) provided a comprehensive analysis of response times and their treatment.

Probability. Interpreted as a relative frequency, probability is the number of times at least one event occurred within a context, divided by the number of opportunities for it to occur. If the context is temporal, probabilities are closely related to rates and latencies: The phrase “at least one” stops the count after one, so probability is a kind of rate with a ceiling of 1.0. The number of first responses divided by the number of epochs yields a probability; the number of first responses divided by the time available to make them yields a reciprocal latency; the total number of responses divided by the time available to make them yields a rate. The maps among these variables are drawn in a companion article (Killeen, Hall, Reilly, & Kettle, 2001²).

Persistence. The number of responses emitted in extinction is a classic measure of strength. It was used by Skinner to measure the reflex reserve (despite its conflict with the rate measure of the reserve). The relation between rate and number of responses in ex-

¹ To see this requires placing such measures in the context of the general mean theorem (Hardy, Littlewood, & Polya, 1959). Then the arithmetic mean $M_1 = [\sum f(x)x^1]^1$, where $f(x)$ is the frequency of a particular IRT (x). Thus the frequency is weighted by the value of the IRT. If the IRTs are not binned, this reduces to $M_1 = \sum x/n$. The harmonic mean, $M_1 = [\sum f(x)x^{-1}]^{-1}$, thus weighting the frequencies by their reciprocals. The geometric mean is the limit as r approaches 0, $M_{r \rightarrow 0} = [\sum f(x)x^r]^r$. It is clear that for any x , the limiting value of x^r as r approaches 0 is 1. Thus the frequency of each IRT is weighted by 1. What is less obvious is that the resulting average approaches the geometric mean as r approaches 0. That, however, is easily verified by anyone with a pocket calculator.

² Killeen, P. R., Hall, S. S., Reilly, M. P., & Kettle, L. C. (2001). *Models of response rate, probability, and latency*. Manuscript submitted for publication.

tion is nonlinear: Reinforcements beyond the first have a greater effect on the number of responses in extinction than they do on response rate (Skinner, 1938). Persistence therefore tells us something different than rate (Eisenberger, 1992). Unfortunately, its measurement is usually confounded with rate: An easily made response such as key pecking may occur thousands of times in extinction; yet more laborious responses, emitted at a lower rate, may persist much longer. Therefore, the *rate* of decay in response rate is a reasonable alternative measure of persistence, as is the inverse measure half-life, which is the time required for response rate to decrease by half (Nevin, 1988).

Most important, the extinction operation permits assessment of how all properties and their measures covary as they change. Persistence forms the heart of Nevin's general treatment of strength as momentum (Nevin, 1992; Nevin & Grace, 2000; Nevin, Mandell, & Atak, 1983) to which we return in the general discussion. We do not here attempt to describe how behavior changes in time; but rather use the extinction (and satiation) operations to describe how one property changes with changes in the other properties.

Research Plan

Our strategy is to conduct experiments that permit measurement of the temporal properties and their covariation under various motivational operations, and to use these characterizations to construct a measure of strength. The analyses will be based on data collected from pigeons whose key pecking was reinforced during trials with various low probabilities and then was weakened by satiation or extinction. The experimental manipulations include varied trial durations (Experiments 1 and 3), varied rates of reinforcement (Experiment 1), satiation (Experiments 2 and 3), and variable-interval (VI; Experiment 1), variable-ratio (VR; Experiment 2), and fixed-interval (FI; Experiment 3) schedules of reinforcement. Other subjects and responses are reported in Killeen et al. (2001). Most of the analyses were based on the results from Experiment 1 and were subsequently applied to the data from all experiments.

EXPERIMENT 1: VARIABLE-INTERVAL SCHEDULES IN BRIEF TRIALS

Method

Subjects. Four adult homing pigeons (*Columba livia*) with prior experience under similar experimental conditions were used. They were housed in individual cages with a 12:12 hr light/dark cycle (lights on at 8:00 a.m.). All were maintained at 80% of their free-feeding weights and had free access to grit and water. Supplementary food, consisting of fortified mixed grain, was given at the end of each day to maintain prescribed body weights.

Apparatus. A single LVE® operant conditioning chamber with the front wall (30 cm by 35 cm) containing three circular keys arranged horizontally, 21 cm from the chamber floor and spaced 6.5 cm apart, was used. The center key was a Gerbrands response key (2 cm in diameter) that could be transilluminated with white light. The side keys were standard 2.5-cm plastic keys located 11 cm from the side walls. Responses to the side keys had no scheduled consequences and were not recorded. Background illumination was provided by a houselight located in the top right corner of the front wall. A grain hopper (6 cm by 5 cm) was centrally located below the center key, 6 cm from the chamber floor. When activated, a light mounted in the ceiling of the hopper opening illuminated the hopper. Programs written in Microsoft® Quick-Basic 4.5 controlled and recorded all experimental events on an IBM-compatible 386 DX2 computer that was located atop the sound-attenuating chamber. White noise was delivered into the chamber through a small speaker, and additional masking noise was provided by a ventilation fan mounted on the wall opposite the interface panel, for a combined ambient noise level of approximately 72 dB.

Procedure

Trials. Experimental sessions were conducted 6 days per week. Prior to each trial, the chamber was dark for 9 s. These periods were followed by a 1-s warning stimulus, during which the side keys were illuminated red. Trials began with the illumination of the center response key and the houselight. Trials were

Table 1
Sequence of schedules in Phase 2 of Experiment 1. VI values are in seconds.

Bird 50	Bird 93	Bird 94	Bird 95	Sessions
VI 480	VI 960	VI 240	VI 120	8
Extinction	Extinction	Extinction	Extinction	4
VI 120	VI 240	VI 960	VI 480	6
Extinction	Extinction	Extinction	Extinction	6
VI 240	VI 480	VI 120	VI 960	12
Extinction	Extinction	Extinction	Extinction	6
VI 960	VI 120	VI 480	VI 240	14
Extinction	Extinction	Extinction	Extinction	6

scheduled to last for an epoch, often 10 s, during which responses were reinforced according to a 10-interval constant-probability VI schedule. Trials were terminated after reinforcement. Reinforcers that were scheduled but not delivered by the end of a trial remained available for delivery at the beginning of the next trial, contingent on responding. During delivery of a reinforcer, the center response key was darkened and the hopper was activated to provide access to milo grain.

Pretraining. Reinforcers were available for effective responses according to a constant-probability VI 20-s schedule. This condition lasted for 15 sessions. On even-numbered sessions, trials were scheduled to last 10 s. On odd-numbered sessions, trials were scheduled to last 2, 4, 8, 13, or 23 s, with the different trial lengths presented in randomized order. Hopper duration was 2.9 s, and sessions terminated after 60 reinforcers had been delivered.

Phase 1. Each subject experienced a progression of VI schedules: VI 60 s (15 sessions), VI 120 s (27 sessions), VI 240 s (15 sessions), VI 480 s (15 sessions), VI 960 s (18 sessions), and VI 2,000 s (28 sessions). Within each condition, sessions containing fixed-duration trials were alternated with sessions containing variable-duration trials scheduled to last 2, 4, 8, 13, or 23 s. For the last condition (VI 2,000 s), a single reinforcer was randomly presented during a session. Hopper duration was 3.2 s, and sessions terminated after 200 trials.

Phase 2. Each pigeon was reexposed to conditions VI 120 s, VI 240 s, VI 480 s, and VI 960 s, with trial duration constant at 10 s. Hopper duration was 3.2 s, and sessions ter-

minated after 200 trials. Blocks of VI sessions alternated with blocks of sessions containing no reinforcers. These latter extinction sessions were identical to the conditioning sessions except for the omission of reinforcement, and trials therefore always lasted the full 10 s. Table 1 shows the order of conditions for each bird.

Data analysis. All analyses were conducted using the last five sessions of data from each reinforcement condition—five sessions containing fixed-duration trials and five sessions containing variable-duration trials—and all of the data from the extinction trials.

The *probability* of responding on a trial is the ratio of the number of trials during which at least one response occurred over the total number of trials in a session (200). The *latency* of the first response is the total time that elapsed before the first response (on trials during which responding occurred), divided by the number of such trials. Trials without a response were omitted from the latency analysis. Latencies were then converted into the proportion of a trial spent responding by dividing by the trial length (10 s) and subtracting from 1: $1 - L/T$, where L is the average latency and T is trial length. This linear transformation provides a normalized measure that varies in the same direction as probability and rate under manipulation of independent variables. It is more directly conceptualized as a component of response rate than latency per se (see Equation 1 below). *Running rate* is the reciprocal of the average IRT after the first response: $(n - 1)/(T - L)$, where n is the number of responses. Trials without at least two responses were omitted from the running-rate analysis. The *overall rate* is the total number of responses during

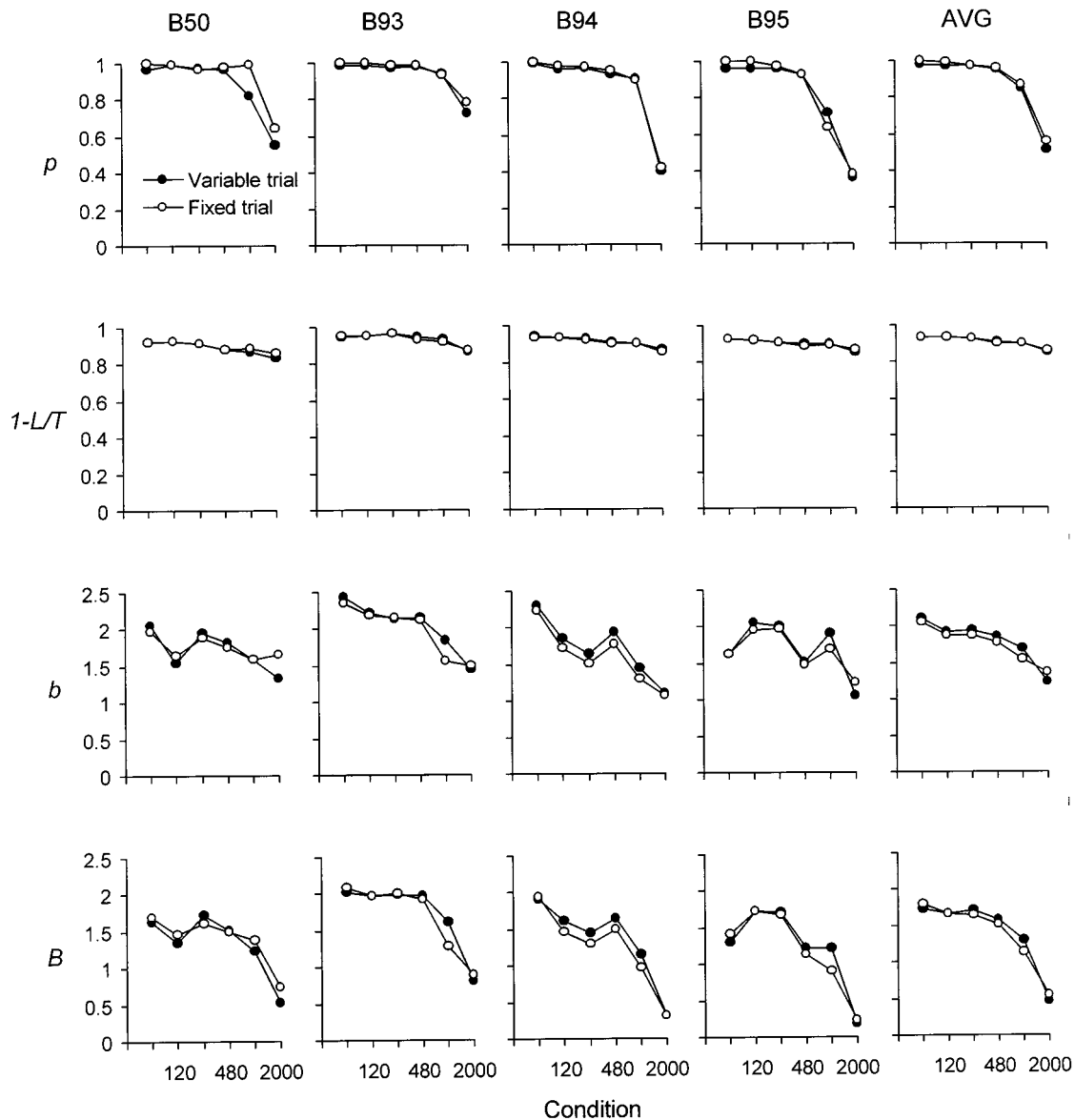


Fig. 1. Changes in the dependent variables for the 4 pigeons (and for the mean) in Phase 1 of Experiment 1 as a function of the VI schedules operative during the trials, and whether the trial durations lasted for exactly 10 s (open symbols) or varied with a mean duration of 10 s (filled symbols). The probability of responding on any trial (p) remained near its ceiling except for the longest VI schedules. The proportion of a trial in the response state ($1 - L/T$) is the complement of the latency relative to the average trial duration; it showed a small but systematic decrease as a function of the schedule. Running rate (rate after the first response, b) and overall rate (total number of responses divided by the number of seconds available for responding, B) decreased more substantially as a function of schedule value.

a session divided by the total time available for responding.

Results and Discussion

Figure 1 shows the dynamic changes in the key variables under the different VI schedules

in Phase 1, for the 10-s trials and averaged over all trial lengths for the variable-duration trials. Performances of the variables are similar across trial type and vary systematically in the predicted directions with changes in the rate of reinforcement. Data sorted by trial

Table 2

Pearson product-moment intercorrelations of measures from Experiments 1 and 2.

	p	$1 - L/T$	b
Experiment 1, reinforcer rate ($n = 12$)			
$1 - L/T$.948		
b	.924	.952	
B	.982	.968	.970
Experiment 1, extinction ($n = 28$)			
$1 - L/T$.871		
b	.852	.907	
B	.967	.904	.918
Experiment 2, satiation ($n = 36$)			
$1 - L/T$.935		
b	.850	.892	
B	.973	.912	.915
Experiment 2, extinction ($n = 28$)			
$1 - L/T$.567		
b	.282	.387	
B	.960	.630	.391

length were unfortunately lost before analyses.

There are many models of relations between independent and dependent variables (e.g., Killeen, 1998); here, the focus is on the relations among the dependent variables. The correlations among the static properties were high, as shown in Table 2. These variables are clearly measuring the same thing, which we call *strength*. They are assimilated to that construct in a factor analysis.

Figure 2 shows cumulative records across sessions of extinction from Phase 2, with the various baseline schedules represented parametrically. After the first condition only four sessions of extinction were conducted, and six thereafter. A smooth and prolonged extinction process over sessions is visible, with ripples, especially noticeable during the last two sessions, indicating some spontaneous recovery. The slope near the origin provides an estimate of the response rate on the base-

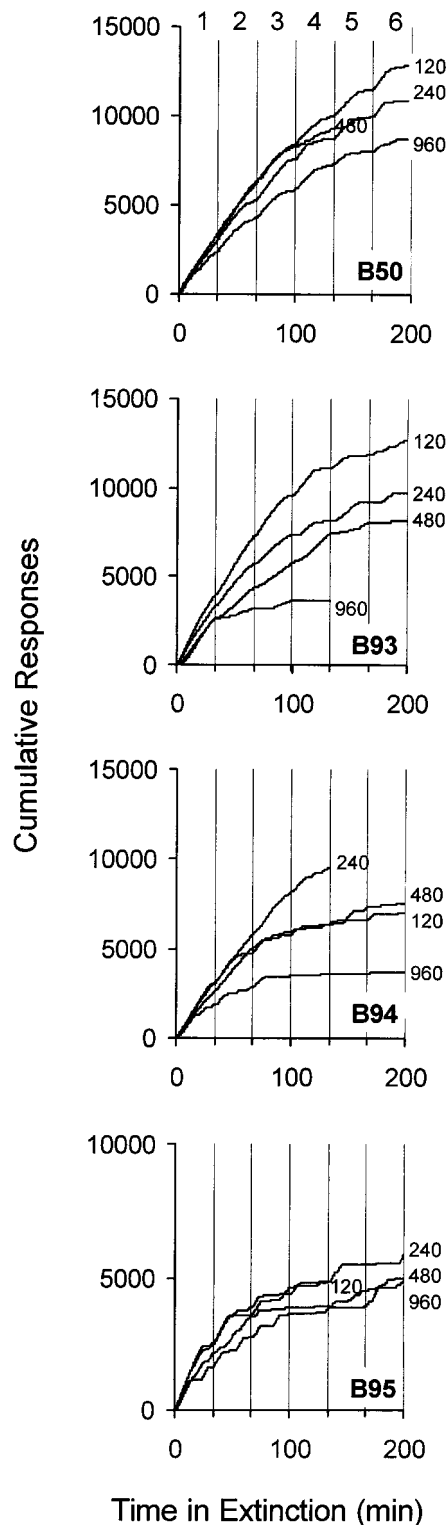


Fig. 2. The cumulative number of responses in extinction after each of the Phase 2 baseline conditions for each of the subjects, with VI value as a parameter. The vertical lines mark the terminations of each session. Under the first condition for each subject, only four sessions of extinction were conducted. The total number of responses in extinction tends to be greatest after schedules with higher probabilities of reinforcement on a trial (120,

240) than after schedules with lower probabilities of reinforcement (480, 960).

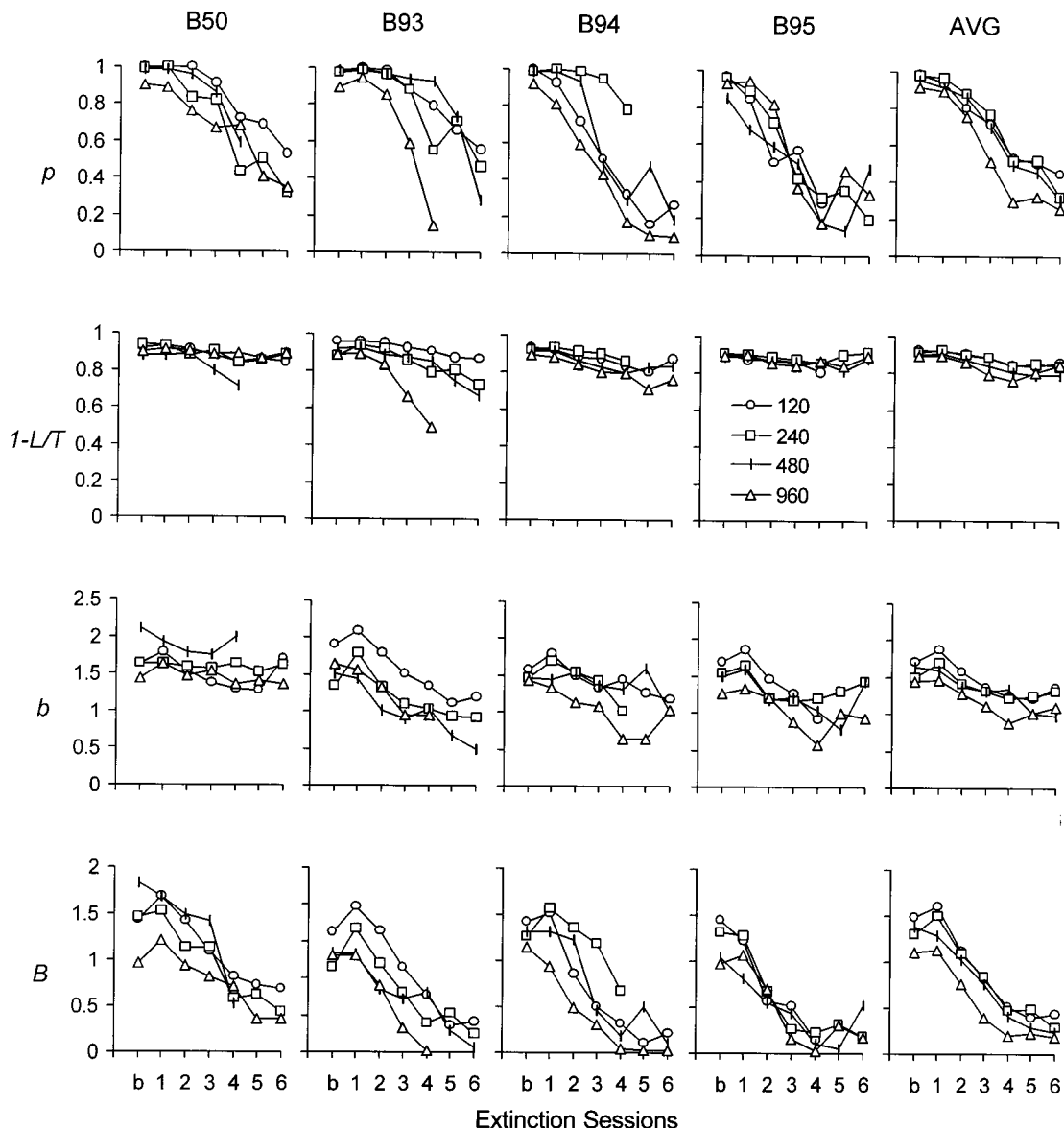


Fig. 3. Changes in the dependent variables as a function of baseline condition (signified by b on the x axis) and subsequent number of sessions in extinction for each subject in Phase 2 of Experiment 1. The probability of responding on any trial decreased throughout the extinction process, as did run rate, b , and overall rate, B . Latency (time to the first response given that one occurred) increased with extinction, whereas its complement decreased.

line schedule. Typically a greater number of responses were emitted in extinction following a richer schedule of reinforcement. Although these cumulative records provide a good synopsis of the effects of withholding reinforcement, it is difficult to draw more precise inferences from them (Killeen, 1985b).

Figure 3 gives the corresponding proba-

bilities, latencies, run rates, and overall rates of key pecking calculated in each condition for each bird at baseline and across sessions in extinction following baseline conditions from Phase 2. The baseline schedules provided reinforcers with probabilities ranging from 8% to 1%. Over this range the independent variable had consistent effects on the dependent variables

(first points in each panel), which are best discerned in the rightmost column, averaged across subjects. The probability of responding at all baseline schedules except the leanest is close to 1 (average .97), with the average probability for the VI 960-s schedule being lowest at .91. These high probabilities are impressive, given that the probability of reinforcement on any particular trial of the VI 960-s condition was only about 1% (two reinforcements per session).

Latencies varied as a function of both baseline probability of reinforcement and extinction session. As noted above, these are latencies *given that a response occurred*. They therefore underestimate the effects of both baseline conditions and extinction sessions: As the probability of responding decreased across conditions and sessions, there was an increasing probability that very long latencies were terminated by the end of the trial, and thus were lost to the sample. These could be appropriately captured by assigning an arbitrarily large latency to trials with no response and then calculating median latencies. That was not done here.

On average, run rates were highest for the VI 120-s condition and lowest for the VI 960-s condition, and decreased with sessions of extinction. The bottom row of Figure 3 combines the information in the first three rows, plotting the overall response rate, calculated as number of responses during a session divided by the time available for responding; that is, as response rate given trial stimulus.

The rightmost columns of Figures 1 and 3 portray the averages across subjects. The baseline schedules had an effect on all the variables, and the changes in those variables were congruent, given those different starting points determined by the baseline schedules. The harmony of measures under Phases 1 and 2 is shown even more clearly by the intercorrelations in Table 2. This table gives the Pearson product-moment correlations of the three fundamental measures and their composite, overall response rate. These are uniformly high, with the strongest relation being between overall response rate and the probability of responding on a trial. The constellation of measures is strongly correlated with the baseline reinforcement schedules and the number of sessions of extinction. Hearst

(1961) reported similar dependencies. Because of the way they were measured, probability, latency, and run rate are theoretically independent of one another. Latency might be a more sensitive measure if trials without a response were counted as long latencies. Overall response rate is a composite of the individual measures, so its high correlation with them is foreordained.

EXPERIMENT 2: VARIABLE-RATIO SCHEDULES

To determine whether these results are idiosyncratic to VI schedules, 4 new pigeons were exposed to VR schedules in which only 1 of every 20 responses would, on the average, be reinforced. The variables were manipulated by letting the animals become satiated through the course of the session, and by experimental extinction.

Method

Subjects. Four adult homing pigeons (*Columba livia*) with prior experience under similar experimental conditions were used. They were housed in individual cages with a 12:12 hr light/dark cycle (lights on at 7:00 a.m.). All were maintained at approximately 84% of their free-feeding weights and had free access to grit and water. Fortified mixed grain was given at the end of the day to maintain prescribed body weights.

Apparatus. An LVE® operant conditioning chamber contained a front wall (30 cm by 35 cm) with two keys (2.5 cm diameter) arranged horizontally on the panel, 21 cm from the chamber floor and spaced 12 cm apart. Responses to the second key had no scheduled consequences and were not recorded. A houselight, located in the top center of the front wall, provided general illumination. A grain hopper (6 cm by 5 cm) was centrally located, 6 cm from the chamber floor. When activated, a light mounted in the top of the hopper opening illuminated the hopper. Programs written in Microsoft® Quick-Basic 4.5 controlled and recorded all experimental events from a computer located atop the sound-attenuating chamber. A speaker delivered white noise into the chamber, with additional masking from a ventilation fan, yielding an ambient noise level of approximately 72 dB.

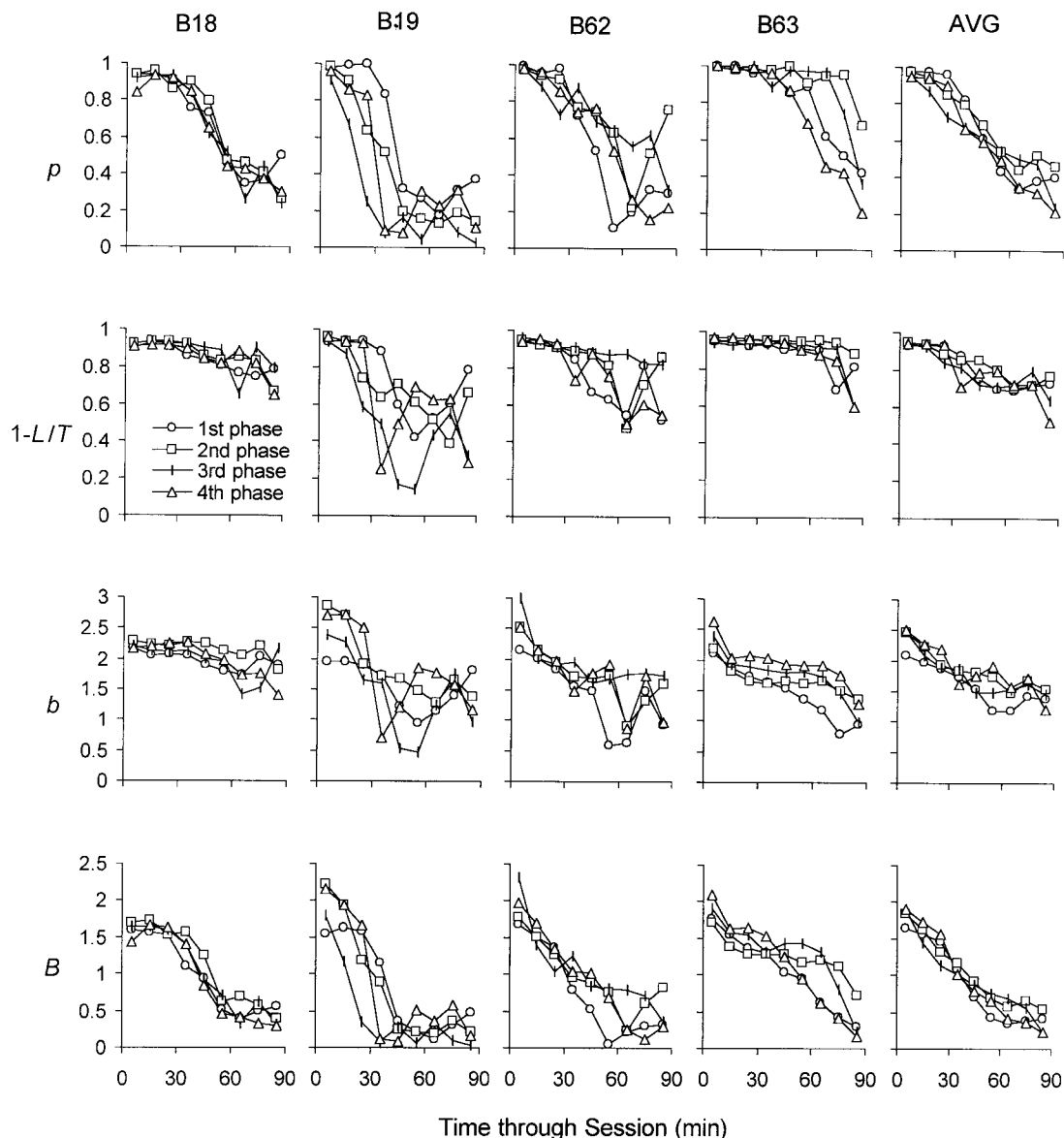


Fig. 4. Changes in the dependent variables as a function of satiation through the course of a session for pigeons whose key pecking was reinforced according to VR schedules in Experiment 2. The various phases are replications of the same procedure.

Trials. Experimental sessions were conducted on alternate days, 3 days per week. Session lasted for 2 hr, giving the subjects exposure to 360 trials. Prior to each trial the chamber was dark for 9 s, followed by a 1-s warning stimulus during which the houselight flickered. Trials began with the illumination of the houselight and the left response key with green light (B18 and B62) or the right re-

sponse key with red light (B19 and B63). Trials were scheduled to last for 10 s, during which responses were reinforced with a probability of 1/20. Trials were terminated after reinforcement. During delivery of a reinforcer, the response key was darkened and the hopper was activated to provide 3.5-s access to milo grain. This condition lasted for 10 sessions, followed by one session of extinction.

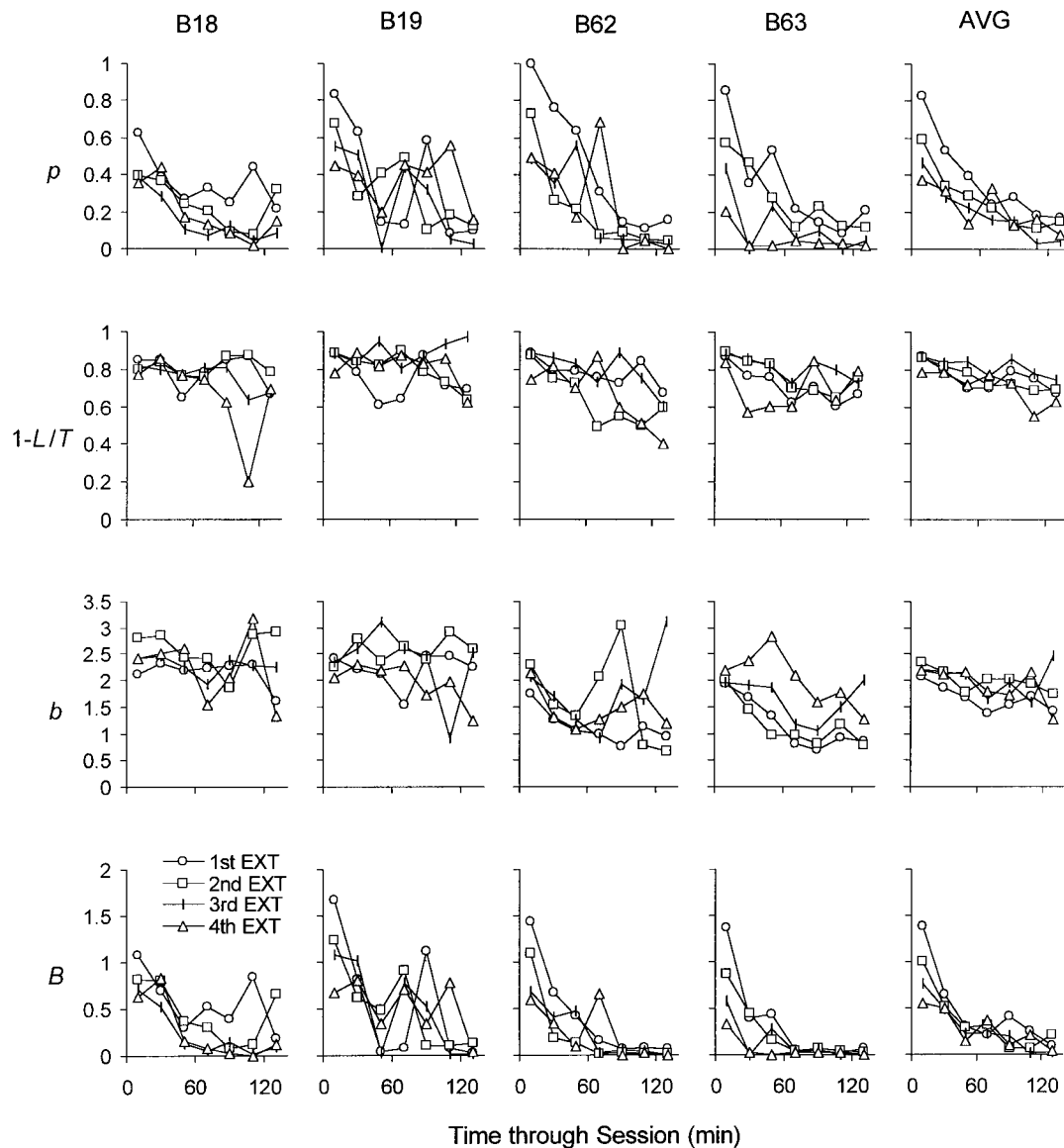


Fig. 5. Changes in the dependent variables as a function of extinction for pigeons whose pecking was reinforced according to VR schedules. The various phases are replications of the same procedure.

This cycle of 11 sessions (a phase) was replicated three additional times. Data from the last five sessions of each phase are reported.

Results and Discussion

Figure 4 shows that the key temporal properties of responding decreased with satiation in a manner similar to that shown in Figures 1 and 3 for VI schedules supporting different rates of reinforcement and undergoing experimental extinction. There are no notable

differences attributable to repeated episodes of conditioning and extinction. Similar data for extinction are shown in Figure 5. Not so obvious in this figure is the dependence of total responses in extinction on the number of prior extinctions: Three of the 4 pigeons emitted more responses in the first extinction phase than in any of the others, and for 2 pigeons the number decreased monotonically with extinction phase.

The intercorrelations among the static

properties during satiation were similar to those found in Experiment 1, and are shown in Table 2. The probability of such high correlations occurring by chance in both experiments is remote. Three different operations (varied rate of reinforcement, satiation, and extinction on interval and ratio schedules) are affecting some thing in a similar manner, and each of these three dependent variables is measuring aspects of it.

EXPERIMENT 3: FIXED-INTERVAL SCHEDULES

In an attempt to extend the generality of these relations, FI schedules were used in place of VR schedules. If the opportunity to collect a reinforcer ceases soon after the end of the trial, it is called a limited hold (LH) technique, and it preserves the trials structure of the experiment. If the hold is unlimited, then the procedure is more akin to a traditional FI schedule. Both variants were tested in this experiment.

Method

The subjects, apparatus, and basic procedure were the same as in Experiment 2. Sessions lasted for 2 hr, and were conducted on alternate days, allowing birds to become satiated during the course of the session and return to a running weight of approximately 84% ad libitum by the beginning of the next session.

Fixed-interval limited hold. The subjects were trained on an FI 20-s LH 10-s schedule for 15 sessions: The first response to the left key after 20 s had elapsed from trial onset was reinforced; if no response had occurred by 30 s from trial onset, the trial was ended and the intertrial interval of 10 s was initiated. On the 16th session the hopper was empty.

Fixed interval. The pigeons were retrained on FI 20 s for 10 sessions. In this condition, trials continued until a reinforcer was collected, or the session terminated after 120 min. On the 11th session the hopper was empty.

Results and Discussion

Figure 6 shows the changes in the temporal properties as the pigeons became satiated under both conditions. For the condition without the limited hold, the probability of re-

sponding stayed near 1.0 because all but the very last trial *had* to contain a response; the LH condition showed the expected decrease in probability of responding as a function of time through the session. The proportion of the trial spent responding ($1 - L/T$) was calculated using the data for L and T on a trial-by-trial basis and averaged. This measure decreased in a parallel manner in both conditions, with the LH condition generally higher. This is because latencies could not exceed 30 s in the LH condition but could be as long as 2 hr in the condition without the LH. There were individual differences in performance, but average run rates and average overall rates were similar across conditions. All measures except response probabilities in the standard FI condition were highly inter-correlated (see Table 3).

In FI extinction (Figure 7), the probability of a response on a trial remained close to 1.0, because no trial except the last terminated before a response occurred. Latencies increased (and thus the proportion of a trial spent responding decreased) in extinction. Given the fixed session duration and the absence of a limited hold, there were few trial onsets toward the end of the session, and therefore the data points after 60 min of extinction are unreliable.

In the LH condition, given that a bird responded on a trial, the latency did not change very much over the course of extinction: If there was enough strength to respond at all, the response tended to occur about one third of the way into the trial (although B88 showed a regular trend in this measure).

These measures are by their nature complementary; in the LH condition latencies greater than 30 s did not affect the latency measure, but decreased the number of trials with a response; in the condition without the LH they did not affect the probability of responding on a trial, but increased the latency.

In all cases, run rate and overall rate decreased with time through the extinction session.

GENERAL DISCUSSION

The measurement of rate. Blough (1963) displayed the first complete pictures of IRT distributions derived from pigeons' pecking. He noted a band of high-probability responses

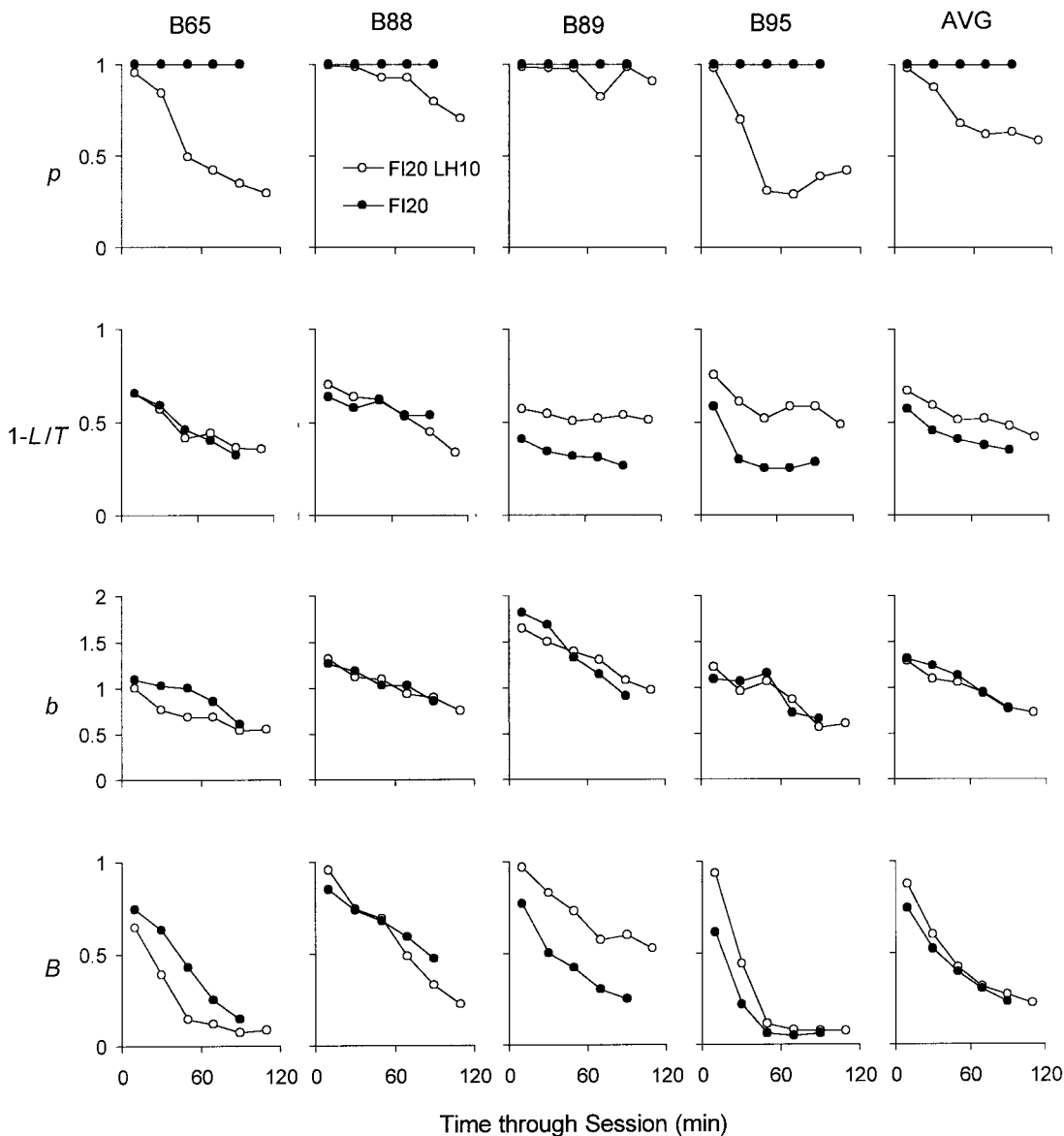


Fig. 6. Changes in the dependent variables as a function of satiation through the course of a session for pigeons whose pecking was reinforced according to FI schedules in Experiment 3. Open symbols are for data collected with an LH contingency; filled circles are for data collected without the LH.

occurring at around 3 Hz whose location did not change under experimental manipulations. The same band was reported by Palya (1991), who also found it resistant to change under schedule manipulation. Blough concluded that responses in this band were primarily under control of proprioceptive feedback from the preceding response: This insensitive "component may have to be given its due, particularly where rate enters into

quantitative relationships" (Blough, 1963, p. 246).

We found that the geometric and harmonic averages of IRTs were almost completely insensitive to changes in the probability of reinforcement, even through the process of extinction. This is because those transformations give too much weight to the minimal IRT, of duration δ . Differences in δ say more about the operandum than about the oper-

Table 3

Pearson product-moment intercorrelations of measures from Experiment 3.

	p	$1 - L/T$	b
Unlimited hold, satiation ($n = 5$)			
$1 - L/T$			
b		.876	
B		.993	.922
Unlimited hold, extinction ($n = 5$)			
$1 - L/T$			
b		.860	
B		.776	.976
Limited hold, satiation ($n = 6$)			
$1 - L/T$.946		
b	.884	.949	
B	.980	.965	.938
Limited hold, extinction ($n = 6$)			
$1 - L/T$.690		
b	.959	.715	
B	.954	.695	.983

ant. Others have noted the “hard-wired” aspect of the fundamental frequency, $1/\delta$ (e.g., Premack, 1965). Crites, Harris, Rosenquist, and Thomas (1967) found good stimulus generalization gradients for rats when the dependent variable was mean IRT but found flat gradients for the median or modal IRT—measures that give more emphasis to the insensitive base frequency, $1/\delta$ (cf. White, 1973). This characteristic frequency is thus paradoxically both a fundamental factor in response rate and at the same time the least interesting aspect of rate. Averaging IRTs and inverting them to a rate gives the largest weight to the longest IRTs. These long IRTs will often correspond to breaks in responding. Thus, the traditional measure of rate gives information about the proportion of time on task. A response-state analysis of rate is provided by Killeen et al. (2001), in which the relation between rate and probability is clarified. In the present paper, the standard measure of rate is used as a molar dependent variable.

Overall response rate. Skinner sought a measure of strength that “describes the state of the reflex with respect to all its static properties at once” (1938, p. 15). Overall response rate (B) does this. It combines the probability of responding on a trial, p , the proportion of the trial spent responding, $1 - L/T$, and the average running rate, b :

$$B = p(1 - L/T)b, \quad L < T. \quad (1)$$

Overall response rate therefore provides a summary statistic indicating the state of all these static variables at once, providing that they are positively intercorrelated. Tables 2 and 3 gave the correlation matrix for p , $1 - L/T$, and b , and also the correlation of these components with their overall composite, B . For individual subjects in Phase 2 of Experiment 1, 23 of the 24 intercorrelations are positive. The strongest correlation is between the probability of responding on a trial and the overall response rate.

Whereas correlational analyses report what each constituent by itself contributes to overall rate, regression analyses report what each contributes in the context of the others. To permit the use of linear regression, take the logarithm of Equation 1:

$$\begin{aligned} \log(B) &= \log(p) + \log(1 - L/T) \\ &\quad + \log(b). \end{aligned} \quad (2)$$

Next, multiply each factor in the right side by a weight, β_i , specified by a regression. As must be the case, the composite is predicted by its elements: Equation 2 accounts for most of the variance in the log rates from these experiments. The β weights in Equation 2 tell how much each element matters in the composition. They show that, for data averaged across all subjects in both conditions of Experiments 1 and 2, probability was the most important element in these experiments (average standardized $\beta_1 = 0.85$) followed by running rates ($\beta_3 = 0.22$). The weight for latency was close to 0.0. The weights will vary depending on experimental conditions: For long trials, for instance, latency of the first response will often play a negligible role in predicting overall rates, in which case its β weight will approach zero. Conversely, on fixed-ratio schedules latency will be a dominant element, with run rates relatively high and constant. This distinction is shown most graphically in the results from Experiment 3, in which different experimental paradigms rendered some indexes so constant as to be mute. This context dependence in the sensitivity of the indicators does not, however, entail a difference in the state of the thing indicated, as shown by the following analyses.

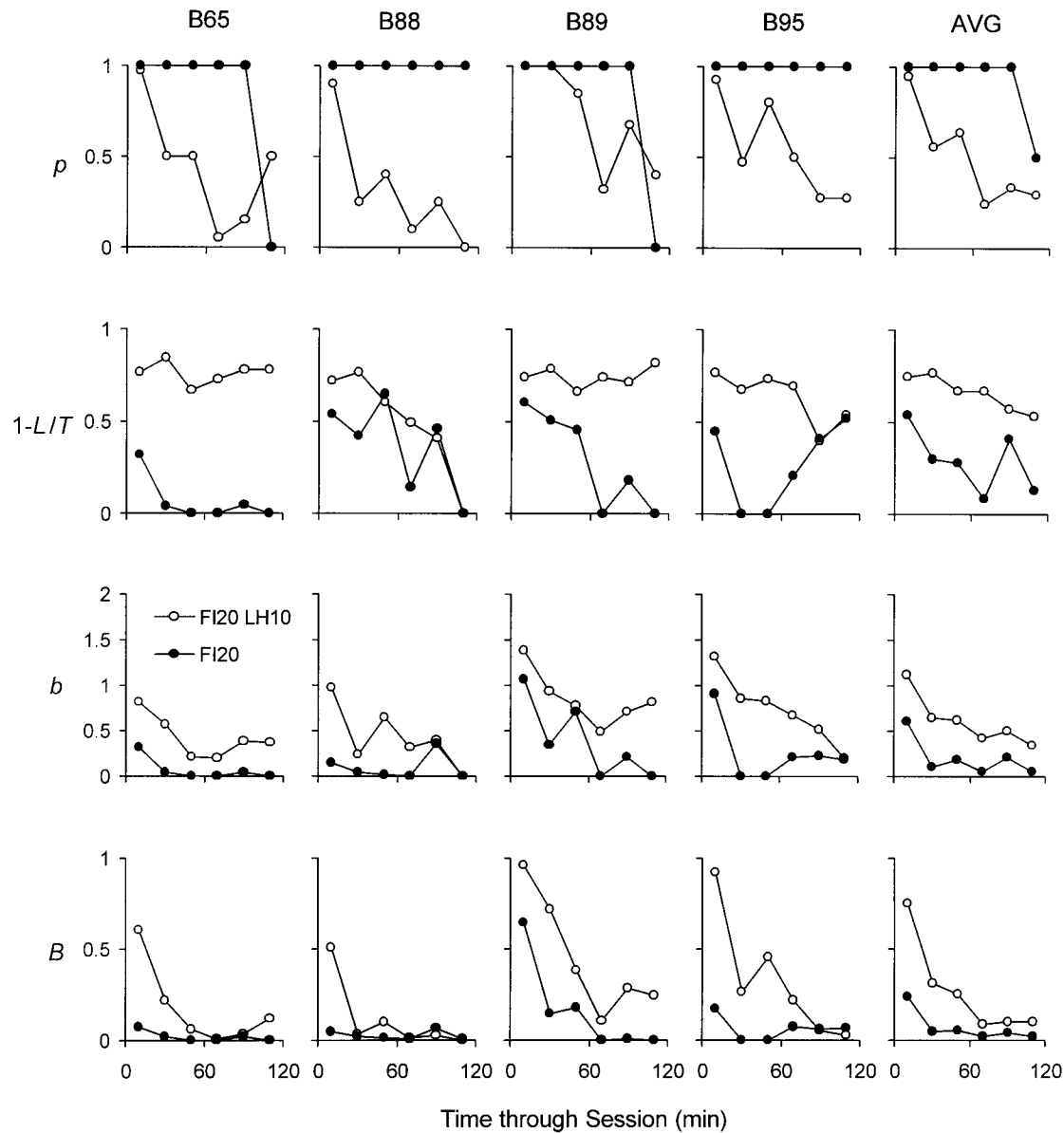


Fig. 7. Changes in the dependent variables as a function of extinction through the course of a session for pigeons whose pecking was reinforced according to FI schedules. Open symbols are for data collected with an LH contingency; filled circles are for data collected without the LH.

Strength

Analysis is "The separation of an intellectual or substantial whole into its constituent parts for individual study . . . to determine either their nature (qualitative analysis) or their proportions (quantitative analysis)" (*American Heritage Electronic Dictionary*, 1992). Qualitative analysis led to the inclusion of some properties of a response and the elim-

ination of others (e.g., amplitude, force). Equations 1 and 2 epitomize a quantitative analysis of response rate. The static properties of probability, latency, and run rate are correlated with one another and with their compendium, overall response rate.

Correlation, as is well known, does not entail causation. It is unlikely that response probability changed *because* run rate changed.

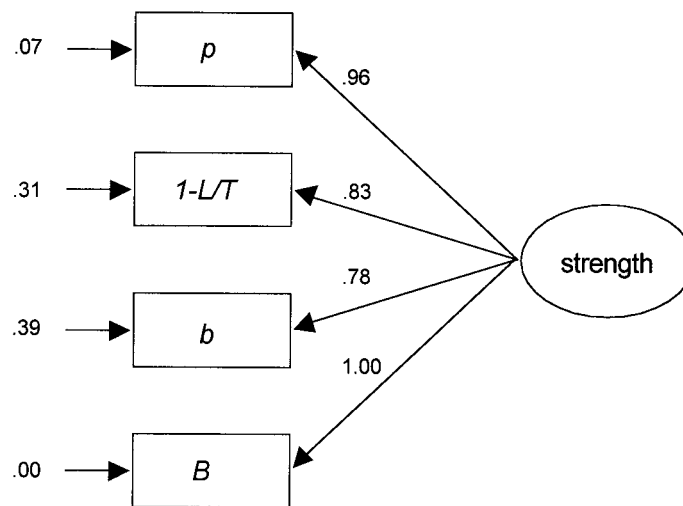


Fig. 8. Results of confirmatory factor analyses, showing the correlations of the hypothetical construct of strength with the dependent variables, based on the data reported in Table 2. The numbers at the left indicate error coefficients.

What is the common causal factor underlying the correlations shown in Table 2? A picture of the common factor may be drawn using principal component factor analysis, which is a technique for reducing a correlation matrix into fewer dimensions (Everitt, 1996; Gentry, Weiss, & Laties, 1983).

Principal component analysis was conducted to determine whether the correlation matrices of p , b , and $1 - L/T$ could be usefully reduced to fewer factors. In this procedure a line is drawn through the data such that it accounts for the largest amount of variance. This line is called the first principal component. The second principal component is the second line perpendicular to the first that resolves the next largest amount of variance remaining in the data, and so on. For Birds B93, B94, and B95, the first principal component explained 84.4%, 85.5%, and 74.4% of the variance in the data from Experiment 1 (Phase 2). For these birds a single factor, which we call strength, could concisely "stand for" the three dependent variables. For Bird B50 the first principal component accounted for only 47.6% of the data variance, with the second resolving a further 41.5%. Two components of strength were necessary for this subject due to its negative covariation of latency and run rate. The principal component accounts for 93% of the variance in the average data shown in Figure 3. Thus, despite

differences among individual subjects, we can, in general, assert that changes in these temporal measures of responding are manifestations of a common factor, strength.

Factor analyses can also be used as a confirmatory technique, to establish whether or not a specific model is compatible with the observed correlations. Assuming a single underlying factor of strength, we can use confirmatory factor analysis to describe how well the observed indicators of strength (i.e., p , $1 - L/T$, b , and B) actually measure the underlying construct. The factor analysis model can be represented as follows:

$$\begin{aligned} p &= \rho_1 S + u_1, & 1 - L/T &= \rho_2 S + u_2, \\ b &= \rho_3 S + u_3, & B &= \rho_4 S + u_4, \end{aligned}$$

where S is the strength of the operant, ρ_i is the correlation between strength and the observed variables, and u_i represents the amount of variance in the observed variables not accounted for by strength.

Figure 8 gives estimates of the parameters derived from fitting the model to the correlations shown in Table 2. The correlation coefficients, ρ_i , are shown on the arrows leading from the strength factor to each observed variable. The error coefficients, u_i , are shown on the left of each variable. Figure 8 indicates that strength is best estimated by B and p , with factor loadings of 1.00 and .989, respec-

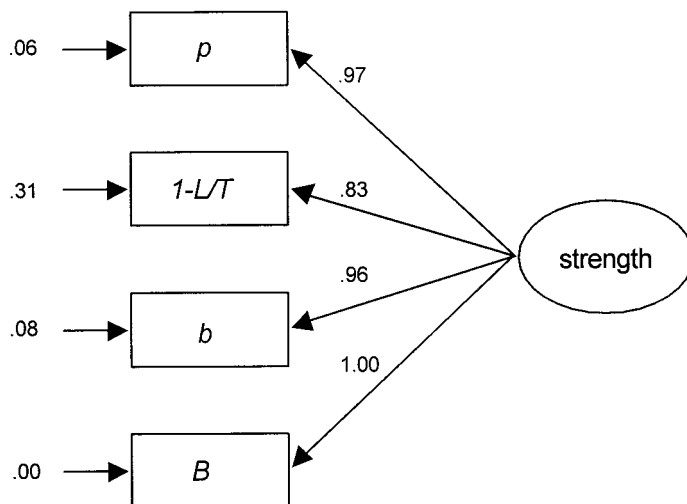


Fig. 9. Results of confirmatory factor analyses showing the correlations of the hypothetical construct of strength with the dependent variables based on the data from Experiment 3 reported in Table 3. The numbers at the left indicate error coefficients.

tively. The error variances corresponding to B and p are not significantly different from zero, suggesting that either variable is a satisfactory measure strength of responding.

The square of the estimated loading of a variable on the factor can be considered to be an estimate of its reliability (see Everitt, 1996). The reliabilities of p , $1 - L/T$, b , and B are therefore .978, .741, .797, and 1.0, respectively. These statistics are for the data averaged over subjects, and are lower for individual birds.

A similar analysis was conducted on the data from Experiment 3, for the LH conditions involving both satiation and extinction (see Table 3). The results are shown in Figure 9. With the exception of a higher relation between run rate and strength, the results for the two analyses (Figures 8 and 9) are virtually identical.

The common factor of strength thus accounts for most of the variance in run rate, latency, and probability of responding. Overall response rate, a traditional measure of strength (Rachlin, 1994), loads heavily on that dimension. S. Roberts (1987) cited Mackintosh: "We do not, in general, have any independently validated way of transforming our arbitrary records of speed, amplitude, probability or rate of responding into more meaningful measures of underlying associative value, response strength, or reaction po-

tential" (1974, p. 494). But Roberts's "multiplicative factors method" showed that "somewhat the opposite is true: A common response scale—response rate—reveals structure not predicted by the usual theories" (1987, p. 158). Our analyses have been concerned with the internal validity of response rate, and Roberts's is concerned with its external validity; however, our conclusions are the same.

Reification. The behavioral community has been justifiably chary of hidden causes in theoretical accounts of behavior. How many things must covary before it is proper to reify a common factor as a mediating cause? Certainly more than two. Attributing the status of "reinforcer" to an event that increases response rate in only one context is fatally circular for a causal account. Historically, it was the transsituational validity of reinforcers that increased the degrees of freedom in their predictive ability beyond that which was absorbed by their nomination. But *reinforcer* is only a binary attribute: An event can only reinforce or fail to, with no gradations. A science of behavior analysis can do better; it can determine the functions of which behavior is a variable. Some events are stronger reinforcers than others are, and their aspects may be rescaled and recombined to predict their reinforcement strength as a continuous function of such relevant variables (e.g., Killeen,

1985a; Killeen, Cate, & Tran, 1993). A similar situation holds for the strength of the operant, of which the reinforcer is but one component.

There are dangers in reification. Gould (1981/1996, chap. 6) astutely criticizes Spearman's reification of the principal component of human intellectual abilities, *g*, as general intelligence. Many behavior analysts question the utility of constructs such as *memory* (e.g., Branch, 1977; Lattal & Abreu-Rodrigues, 1994) and *internal clocks* (Higa, 1999; Zeiler, 1998). But there are benefits to such theoretical commitments (Staddon, 1993). Hypothetical constructs serve as a bridge between related areas of research (Wilkie & Saksida, 1994). They focus intellectual energy, and even imperfect focus can lead to more effective procedures than diffuse manipulation of variables. There is some risk: Resources may be wasted in research that rejects incorrect hypotheses. But even more may be wasted in research that answers no one's questions.

Successes in this constructive approach to science redeem its failures. The gene existed in theory before DNA was discovered, and remains a useful construct. Maxwell's field equations were developed on the foundation of the luminiferous ether, and remain useful. Newton reified the principal component of the movements of pendula and tides and moons as *gravity*, after first transforming those variables according to a theory, and then "mending" his estimates of its force to bring its diverse measures into coincidence with their principal component (Westfall, 1973). Whereas Newton was cautious to frame no hypotheses concerning mechanism—no plausible ones were apparent to him—he was bold enough to overcome his mechanistic training and posit a universal, invisible, "true" cause, gravity.

His critics, however, proceeded in just the opposite manner, starting out with the vexing problem of how such a force as the proposed Newtonian gravitation can possibly exist and act according to Newtonian laws, and not accepting the formal results of the *Principia* so long as they did not find its conceptual basis to be satisfactory. These critics, in other words, were not willing to go along with the procedural mode of the Newtonian style. (Cohen, 1995, p. 143)

Fortunately, it is not necessary for behav-

iorists to abandon a positivistic stance to go along with the results of the present analysis. Global response rate is a good measure of the principal component of strength. Rates can be used with assurance that they are correlated with a fundamental underlying factor, without having to acknowledge that factor.

Reflex strength. Skinner cited the "demonstration" that the rate of response was the principal measure of the strength of an operant (Point 4 in the quotation cited in the introduction). There was no such demonstration. Selection of rate was a Hobson's choice, forced by the loss of other properties that required a predefined epoch (e.g., a trial from whose onset a latency could be measured, and one of sufficiently limited duration to make the probability of the response a useful measure); or that required an alternative response to measure choice; or that required difficult measurements (e.g., of force). Skinner called behavior in trials experiments "pseudo-reflexes." Discriminative stimuli often exert strong control over behavior. Wenrich (1963) noted that discriminated operants are much more resistant to satiation than are free operants. In Experiment 1 pigeons responded on over 90% of the trials when the probability of reinforcement was only 1%. We have seen pigeons respond on over 90% of the trials for many weeks in Pavlovian paradigms in which responses eliminate reinforcement ("negative automaintenance").

To free behavior from the excessive control of eliciting stimuli, Skinner downplayed the role of the discriminative stimulus and emphasized the free operant (Lindsley, 1996), which relegates the stimulus to the experimental context or to the passage of time. This tactic unfortunately discouraged analyses of the continuum between free and discriminated operants and the phrasing of questions such as: What is the relation between Pavlov's (1927) *inhibition of delay* and Skinner's post-reinforcement pause? What is the operant analogue of pseudoconditioning? How does an autoshaped response (Locurto, Terrace, & Gibbon, 1981; Peden, Browne, & Hearst, 1977; Rescorla, 1987) become an operant (Moore, 1973)? How does an operant become a respondent (Breland & Breland, 1961)? The analyses of operant strength can provide common ground for a rapproche-

ment of operant and respondent research on such issues.

Rate is an indicator, not an essence. It is possible to reinforce low response rates (e.g., according to a differential-reinforcement-of-low-rate [DRL] schedule) with high rates of reinforcement. Animals prefer such schedules to others that sustain higher response rates with lower reinforcement rates. This seems to pose a difficulty for a simple map between response rate and strength. But DRL contingencies greatly increase the minimal duration of the reinforced response, so that observed rates of completion of these extended responses may be close to their ceiling. An animal not pressing a switch in such situations actually *is* responding, engaging in the first part of the heterogeneous chain that leads to reinforcement. An animal that does other things for 4 s and then lever presses on a schedule under which only IRTs of 5 s are reinforced has not failed to make the correct response; it was making the stipulated response for the first 4 s, and only made the wrong response during the fraction of the second it pressed the switch. A true measure of rate must include in the numerator those kinds of hypothetical behavior that constitute the first part of such chains.

Consider a pianist about to play a concerto; it would seem that probability of playing, latency to start, and rate of key pressing are strictly governed by contingencies and tell us little about strength, or about one another. Rudolph Serkin appears in the concert hall with probability very close to 1.0, his dramatic pause before playing is not an indication of low strength, and the tempi of key pressing is governed by score and interpretation, not variations in strength. But what of the tyro? More than one novice's stage fright has canceled a show. Orchestras seamlessly extend the prelude for artists whose latency to take the stage is extended, whether by traffic or by trepidation. The pause before starting or between notes is no less a part of the performance than key pressing is, but is easily overlooked because it is not measured as switch closure. The concert is a trial, and maintains higher probabilities of performance than the free operant of piano practice, whose rate and latency are good indicators of its strength. Just as television programs can disrupt practice of low strength, catcalls may ter-

minate a concert performance, and may do so more readily when the pianist has doubts about the quality of the piece or of his execution—that is, when strength is weak.

Most important behavior consists of extended and heterogeneous responses (Rachlin, 1994). Particular reinforcement contingencies can change some of the static properties of component responses, leaving naive interpretations of the dependent variables at their mercy. Whereas the latent variable *strength* is inferred from those variables, manipulation of the variables does not necessarily change strength. The static properties are indicators, symptoms if you would, of operant strength. Bending the needle of a voltmeter does not change the strength of an electric field, nor does stifling a shout reduce alarm. It is the coherence of various indicators that lets us recalibrate the voltmeter or interpret the muffled shout. Factor analysis is one method of bringing these indicators into coherence—for defining the scales on the separate measures necessary to achieve unanimity. In the end, it is understanding of the thing they indicate, the hypothetical construct of strength, that is our ultimate theoretical goal, as it was Skinner's.

Momentum. Nevin's theory of behavioral momentum (e.g., Nevin & Grace, 2000) identifies resistance to disruption as a key measure of behavior, one analogous to our construct of strength. The measure of resistance, a unit of what might be called *behavioral mass*, is defined as the amount of a disrupter necessary to reduce baseline rate by the factor $1/e \approx 0.368$ (Killeen, 2000, Equation 1). If the disrupter is experimental extinction, the mass of an operant is the time necessary for its rate to fall to 37% of baseline rate. This formulation factors response rates out of the definition of mass, and does so intentionally, because response rates may be manipulated by contingencies that have little effect on persistence. This appears to be a completely different path to understanding strength: Absolute response rates, our best predictor of strength, are factored out of the measure of mass. Rates are reintroduced, however, in the measure of behavioral *momentum*, which is the product of baseline rate (velocity) and mass, and which predicts the total number of responses emitted as an animal becomes satiated, undergoes extinction, and so on. In the

multiple extinction phases of Experiment 3, most birds emitted fewer responses in the later phases of extinction, indicating a difference in momentum that was more noticeable than were successive changes in rate or latency.

As response rates decrease during extinction or satiation, they indicate a decrease in operant strength. It is to be expected that behavior will also be more easily disrupted at this point, thus indicating a decrease in mass and its correlate, momentum. Both behavioral mass and strength may thus actually be telling the same story.

Relative strength. The notion of absolute strength is something of a fiction, as behavior is always conditional on the total context of operants that are available to the organism. Even though response probability remained high in the VI 960-s condition of Experiment 1, had the chamber door been left open the pigeons might have left for richer patches. Our data are absolute only in the sense of being derived from an environment in which the alternative operants were of a constant low strength. Control by context is manifest in the important role played by the intertrial interval in autoshaping (Gamzu & Williams, 1973; Locurto et al., 1981) and in delayed discriminations (W. A. Roberts & Kraemer, 1982); context plays a leading role in several theories of conditioning (e.g., R. R. Miller & Matzel, 1988; Wagner, 1981). It is an essential part of the measurement of behavioral mass.

In subsequent analyses of these data, Killean et al. (2001) introduce the construct of a nonresponse state, a fugue from measured behavior that is a major source of the variability in rates. The prevalence of nonresponse states is correlated with distractibility, and thus mass (or, more precisely, its inverse). They may indicate a transition to different behavioral modes (Timberlake, 1994). The introduction of competing operants amplifies the role of such fugues, as choice behavior is sensitive to differential amounts of reinforcement that have little effect on absolute response rates (Neuringer, 1967). This insensitivity occurs because ceilings on response rates limit their dynamic range, making them insensitive at high strengths; and because, absent other diversions, responses of even low strength will be emitted, thus saturating the measure. It is as though each ob-

servational epoch is a trial in a signal-detection task, and a response indicates the animal's prediction that a reinforcer will be available. In the experimental chamber, however, almost all of the payoff is for saying "yes"; high probabilities of responding, and corresponding high response rates in real time, are more a matter of bias in the rate measure rather than sensitivity to the underlying dimension of strength. Introduction of a second operant reduces the bias, making relative rates a more sensitive measure of strength. This sensitivity, and the resulting orderliness of the data, made analyses of the relative rates of two operants a central occupation of experimental analysts in the 1970s and 1980s (e.g., Davison & McCarthy, 1988; Williams, 1988).

More than one operant is usually available to most organisms. It follows that relative strength, not absolute strength, will control most of the variance in behavior, both in the laboratory and on the street. By providing a relatively thorough analysis of absolute strength, we have attempted to ground ensuing discussion of choice between operants of different strength.

Analyses. Collecting data is often more fun than analyzing them. But that which reinforces the experimenter does not necessarily strengthen the field:

Postexperimental transformation of the data to reveal order seems to be a very different enterprise from laboratory manipulation of a variable to demonstrate control; a parameter of an equation . . . seems very remote from actual, ongoing behavior. . . . This is nothing new in the analysis of behavior: . . . the very act of counting lever presses or key pecks puts the experimenter at a distance from the moment-to-moment actions of the organism. Moreover, the calculation of response rate or probability over any sample of time or trials also obscures variations in the pattern of responding. (Nevin, 1984, p. 431)

Nevin emphasized the importance of invariances that may be invisible in the raw data but are manifest in the structure of models and their parametric changes (as did Stevens, 1986, pp. 46 ff.). Loftus (1978) takes the argument one step further:

A dependent variable such as response probability is not intrinsically interesting. Rather, it is only interesting because of what it reflects

about a component of some theory. Getting from the dependent variable back to the theoretical component [e.g., strength, mass, gravity] may well involve a transformation on the dependent variable. (p. 318)

To paraphrase Loftus, response rates and probabilities are discriminative stimuli whose value lies in their effective control of our responses to theoretical propositions. To control our behavior most effectively, it may be necessary to transform the data (Shull, 1991). This requires generic laws to complement the generic nature of an operant: laws, that is, that govern underlying states, their strength, and their relation to the independent and dependent variables.

REFERENCES

- Allan, R. W. (1992). Technologies to reliably transduce the topographical details of pigeons' pecks. *Behavior Research Methods, Instruments, & Computers*, 24, 150–156.
- American heritage electronic dictionary (3rd ed.). (1992). New York: Houghton Mifflin.
- Amsel, A. (1962). Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 69, 306–328.
- Amsel, A., & Rashotte, M. E. (1984). *Mechanisms of adaptive behavior: Clark L. Hull's theoretical papers, with commentary*. New York: Columbia University Press.
- Baum, W. M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior*, 20, 137–153.
- Blough, D. S. (1963). Interresponse time as a function of continuous variables: A new method and some data. *Journal of the Experimental Analysis of Behavior*, 6, 237–246.
- Branch, M. N. (1977). On the role of "memory" in the analysis of behavior. *Journal of the Experimental Analysis of Behavior*, 28, 171–179.
- Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, 16, 681–684.
- Brown, J. S. (1948). Gradients of approach and avoidance responses and their relation to level of motivation. *Journal of Comparative and Physiological Psychology*, 41, 450–465.
- Cohen, I. B. (1995). Newton's method and Newton's style. In I. B. Cohen & R. S. Westfall (Eds.), *Newton* (pp. 126–143). New York: Norton.
- Crites, R. J., Harris, R. T., Rosenquist, H., & Thomas, D. R. (1967). Response patterning during stimulus generalization in the rat. *Journal of the Experimental Analysis of Behavior*, 10, 165–168.
- Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Hillsdale, NJ: Erlbaum.
- Domjan, M., & Galef, B. G. (1983). Biological constraints on instrumental and classical conditioning: Retrospect and prospect. *Animal Learning & Behavior*, 11, 115–161.
- Eisenberger, R. (1992). Learned industriousness. *Psychological Review*, 99, 248–267.
- Everitt, B. S. (1996). *Making sense of statistics in psychology: A second-level course*. Oxford: Oxford University Press.
- Filion, S. C., Fowler, J. M., & Notterman, J. M. (1969). Some effects of simultaneous force-proportional positive and negative reinforcement. *Journal of Experimental Psychology*, 82, 267–271.
- Gamzu, E., & Williams, D. R. (1973). Associative factors underlying the pigeon's keypecking in autoshaping procedures. *Journal of the Experimental Analysis of Behavior*, 19, 225–232.
- Garcia, J., McGowan, B. K., & Green, K. F. (1972). Biological constraints on conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: Vol. 2. Current research and theory* (pp. 3–27). New York: Appleton-Century-Crofts.
- Gentry, G. D., Weiss, B., & Laties, V. G. (1983). The microanalysis of fixed-interval responding. *Journal of the Experimental Analysis of Behavior*, 39, 327–343.
- Gould, S. J. (1996). *The mismeasure of man*. New York: Norton. (Original work published 1981)
- Hardy, G. H., Littlewood, J. E., & Polya, G. (1959). *Inequalities*. Cambridge: Cambridge University Press.
- Hearst, E. (1961). Resistance-to-extinction functions in the single organism. *Journal of the Experimental Analysis of Behavior*, 4, 133–144.
- Herrick, R. (1963). Lever displacement during continuous reinforcement and during a discrimination. *Journal of Comparative and Physiological Psychology*, 56, 700–707.
- Higa, J. J. (1999). Interval timing: Is there a clock? *Behavioural Processes*, 45, 1–2.
- Hull, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.
- Kellicutt, M. H. (1967). Response duration during an operant discrimination. *Journal of Experimental Psychology*, 73, 56–60.
- Killeen, P. R. (1985a). Incentive theory IV: Magnitude of reward. *Journal of the Experimental Analysis of Behavior*, 43, 407–417.
- Killeen, P. R. (1985b). Reflections on a cumulative record. *The Behavior Analyst*, 8, 177–183.
- Killeen, P. R. (1988). The reflex reserve. *Journal of the Experimental Analysis of Behavior*, 50, 319–331.
- Killeen, P. R. (1998). The first principle of reinforcement. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp. 127–156). Mahwah, NJ: Lawrence Erlbaum Associates.
- Killeen, P. R. (2000). A passel of metaphors: "Some old, some new, some borrowed. . . ." *Behavioral and Brain Sciences*, 23, 102–103.
- Killeen, P. R., Cate, H., & Tran, T. (1993). Scaling pigeons' choice of feeds: Bigger is better. *Journal of the Experimental Analysis of Behavior*, 60, 203–217.
- Kimble, G. A. (1961). *Hilgard and Marquis' conditioning and learning* (2nd ed.). New York: Appleton-Century-Crofts.
- Lattal, K. A., & Abreu-Rodriguez, J. (1994). Memories and functional response units. *Behavioral and Brain Sciences*, 17, 143–144.
- Levine, G., & Loesch, R. (1967). Generality of response intensity following nonreinforcement. *Journal of Experimental Psychology*, 75, 97–102.
- Lindsley, O. R. (1996). The four free-operant freedoms. *The Behavior Analyst*, 19, 199–210.
- Locurto, C. M., Terrace, H. S., & Gibbon, J. (Eds.).

- (1981). *Autoshaping and conditioning theory*. New York: Academic Press.
- Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, 6, 312-319.
- Logan, F. A. (1956). A micromolar approach to behavior theory. *Psychological Review*, 63, 63-73.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Mackintosh, N. J. (1974). *The psychology of animal learning*. New York: Academic Press.
- Margulies, S. (1961). Response duration in operant level, regular reinforcement and extinction. *Journal of the Experimental Analysis of Behavior*, 4, 317-322.
- Mattes, S., Ulrich, R., & Miller, J. (1997). Effects of response probability on response force in simple RT. *Quarterly Journal of Experimental Psychology*, 50A, 405-420.
- Meyer, D., Osman, A., Irwin, D., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3-67.
- Miller, N. E. (1959). Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2, pp. 196-292). New York: McGraw-Hill.
- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 51-92). New York: Academic Press.
- Moore, B. W. (1973). The role of directed Pavlovian reactions in simple instrumental learning in the pigeon. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Constraints on learning: Limitations and predispositions* (pp. 159-188). New York: Academic Press.
- Neuringer, A. J. (1967). Effects of reinforcement magnitude on choice and rate of responding. *Journal of the Experimental Analysis of Behavior*, 10, 417-424.
- Nevin, J. A. (1984). Quantitative analysis. *Journal of the Experimental Analysis of Behavior*, 42, 421-434.
- Nevin, J. A. (1988). Behavioral momentum and the partial reinforcement effect. *Psychological Bulletin*, 103, 44-56.
- Nevin, J. A. (1992). An integrative model for the study of behavioral momentum. *Journal of the Experimental Analysis of Behavior*, 57, 301-316.
- Nevin, J. A., & Grace, R. C. (2000). Behavioral momentum and the law of effect. *Behavioral and Brain Sciences*, 23, 73-130.
- Nevin, J. A., Mandell, C., & Atak, J. R. (1983). The analysis of behavioral momentum. *Journal of the Experimental Analysis of Behavior*, 39, 49-59.
- Notterman, J. M., & Block, A. H. (1960). Note on response differentiation during a simple discrimination. *Journal of the Experimental Analysis of Behavior*, 3, 289-291.
- Palya, W. L. (1991). Laser printers as powerful tools for the scientific visualization of behavior. *Behavior Research Methods, Instruments, & Computers*, 23, 277-282.
- Pavlov, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London: Oxford University Press.
- Peden, B. F., Browne, M. P., & Hearst, E. (1977). Persistent approaches to a signal for food despite food omission for approaching. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 377-399.
- Ploog, B. O., & Zeigler, H. P. (1996). Effects of food-pellet size on rate, latency, and topography of auto-shaped key pecks and gapes in pigeons. *Journal of the Experimental Analysis of Behavior*, 65, 21-35.
- Premack, D. (1965). Reinforcement theory. In D. Levine (Ed.), *Nebraska symposium on motivation* (pp. 123-180). Lincoln: University of Nebraska Press.
- Rachlin, H. (1994). *Behavior and mind: The roots of modern psychology*. New York: Oxford University Press.
- Rescorla, R. A. (1987). A Pavlovian analysis of goal-directed behavior. *American Psychologist*, 42, 119-129.
- Roberts, S. (1987). Evidence for distinct serial processes in animals: The multiplicative-factors method. *Animal Learning & Behavior*, 15, 135-173.
- Roberts, W. A., & Kraemer, P. J. (1982). Some observations of the effects of intertrial interval and delay on delayed matching to sample in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 8, 342-353.
- Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review*, 77, 406-418.
- Shull, R. L. (1991). Mathematical description of operant behavior: An introduction. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior* (Vol. 2, pp. 243-282). New York: Elsevier.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Staddon, J. E. R. (1993). The conventional wisdom of behavior analysis. *Journal of the Experimental Analysis of Behavior*, 60, 439-447.
- Stevens, S. S. (1986). *Psychophysics: Introduction to its perceptual, neural, and social prospects* (2nd ed.). New Brunswick, NJ: Transaction.
- Svartdal, F. (1993). Working harder for less: Effect of incentive value on force of instrumental response in humans. *Quarterly Journal of Experimental Psychology*, 46A, 11-34.
- Timberlake, W. (1994). Behavior systems, associationism, and Pavlovian conditioning. *Psychonomic Bulletin & Review*, 1, 405-420.
- Timberlake, W., & Lucas, G. A. (1990). Behavior systems and learning: From misbehavior to general principles. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Instrumental conditioning theory and the impact of constraints on learning* (pp. 237-275). Hillsdale, NJ: Erlbaum.
- Townsend, J. T., & Busemeyer, J. R. (1989). Approach-avoidance: Return to dynamic decision behavior. In C. Izawa (Ed.), *Current issues in cognitive processes* (pp. 107-133). Hillsdale, NJ: Erlbaum.
- Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.
- Wenrich, W. W. (1963). Response strength of an operant under stimulus control with satiated subjects. *Journal of the Experimental Analysis of Behavior*, 6, 247-248.
- Westfall, R. S. (1973). Newton and the fudge factor. *Science*, 179, 751-758.
- White, K. G. (1973). Postdiscrimination generalization as a function of interresponse time. *Animal Learning & Behavior*, 1, 297-301.
- Wilkie, D. M., & Saksida, L. M. (1994). A mathematical theory of reinforcement: An unexpected place to find support for analogical memory coding. *Behavioral and Brain Sciences*, 17, 155-156.
- Williams, B. A. (1988). Reinforcement, choice, and re-

- sponse strength. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (Vol. 2, pp. 167–244). New York: Wiley.
- Zeiler, M. D. (1998). On sundials, springs, and atoms. *Behavioural Processes*, 44, 89–99.
- Zuriff, G. E. (1985). *Behaviorism: A conceptual reconstruction*. New York: Columbia University Press.

Received July 10, 2000

Final acceptance January 24, 2001